



Crystallography on a fault-line

Peter Zwart

Berkeley Center for Structural Biology

<http://bcsb.als.lbl.gov>

Introduction



BERKELEY CENTER FOR STRUCTURAL BIOLOGY

PHYSICAL BIOSCIENCES DIVISION

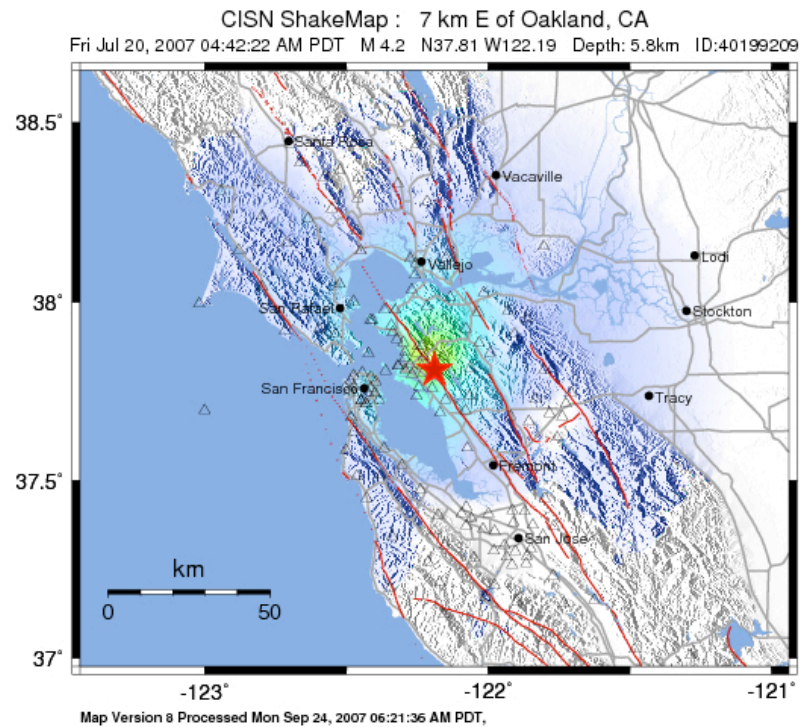
Introduction



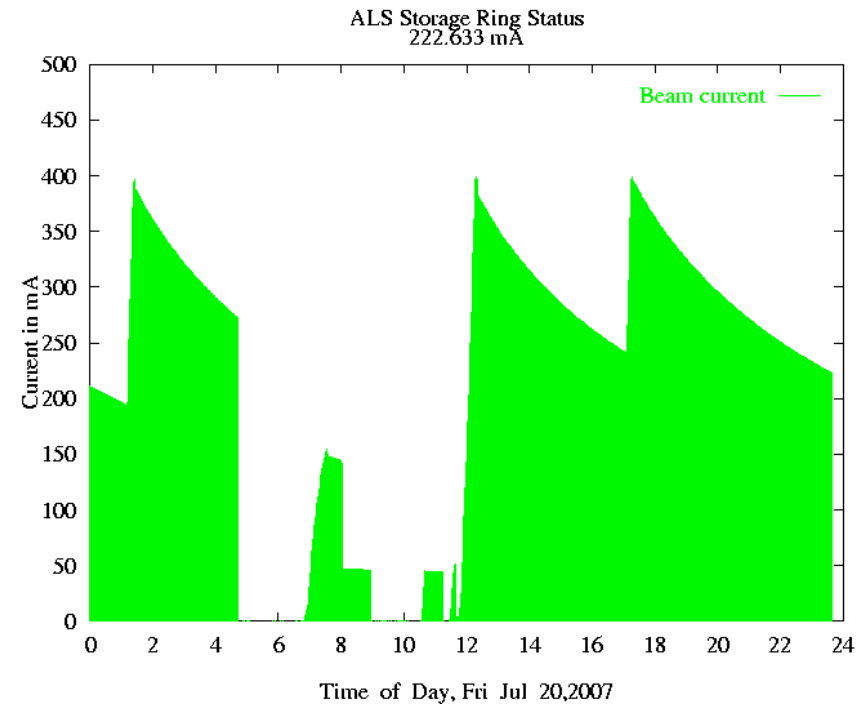
BERKELEY CENTER FOR STRUCTURAL BIOLOGY

PHYSICAL BIOSCIENCES DIVISION

Introduction



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	<.17	.17-1.4	1.4-3.9	3.9-9.2	9.2-18	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	8.1-16	16-31	31-60	60-116	>116
INSTRUMENTAL INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+



BCSB



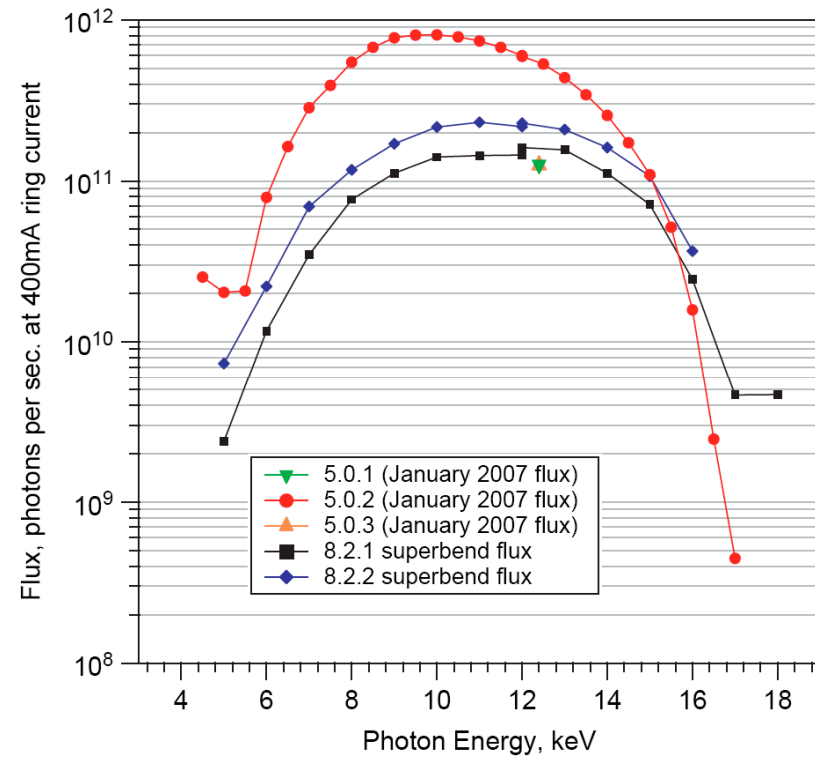
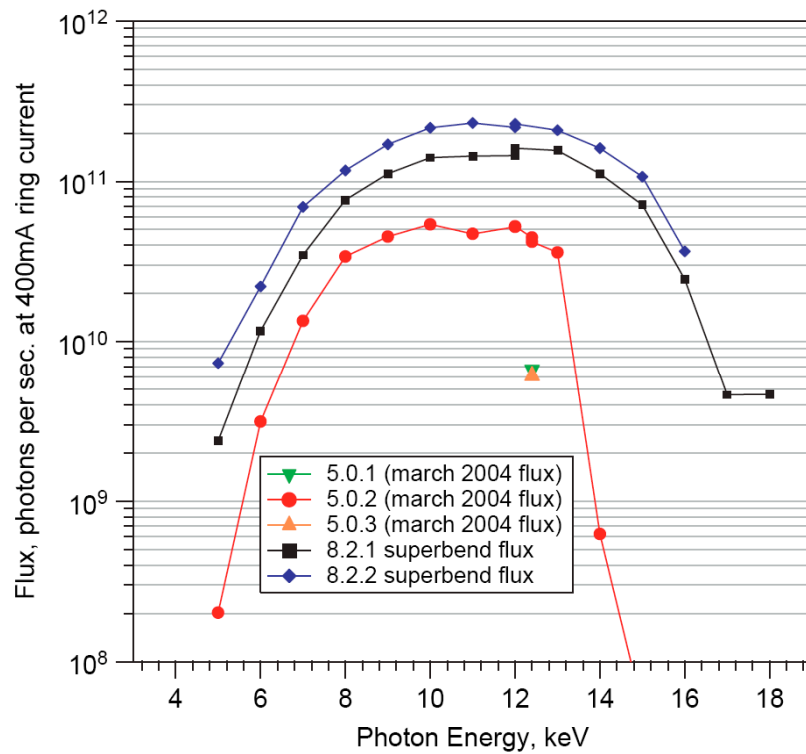
- 5 beam lines
 - 8.2.1, 8.2.2
 - HHMI
 - 5.0.1, 5.0.2
 - Amgen, Vertex, LANL/TBSGX, UCSF, Gilead, Pfizer, FHCRC, Genentech, Celgene, Roche
 - 35% General Users
 - 5.0.3
 - Takeda-SD, GNF
 - 25% General users

Sector 5

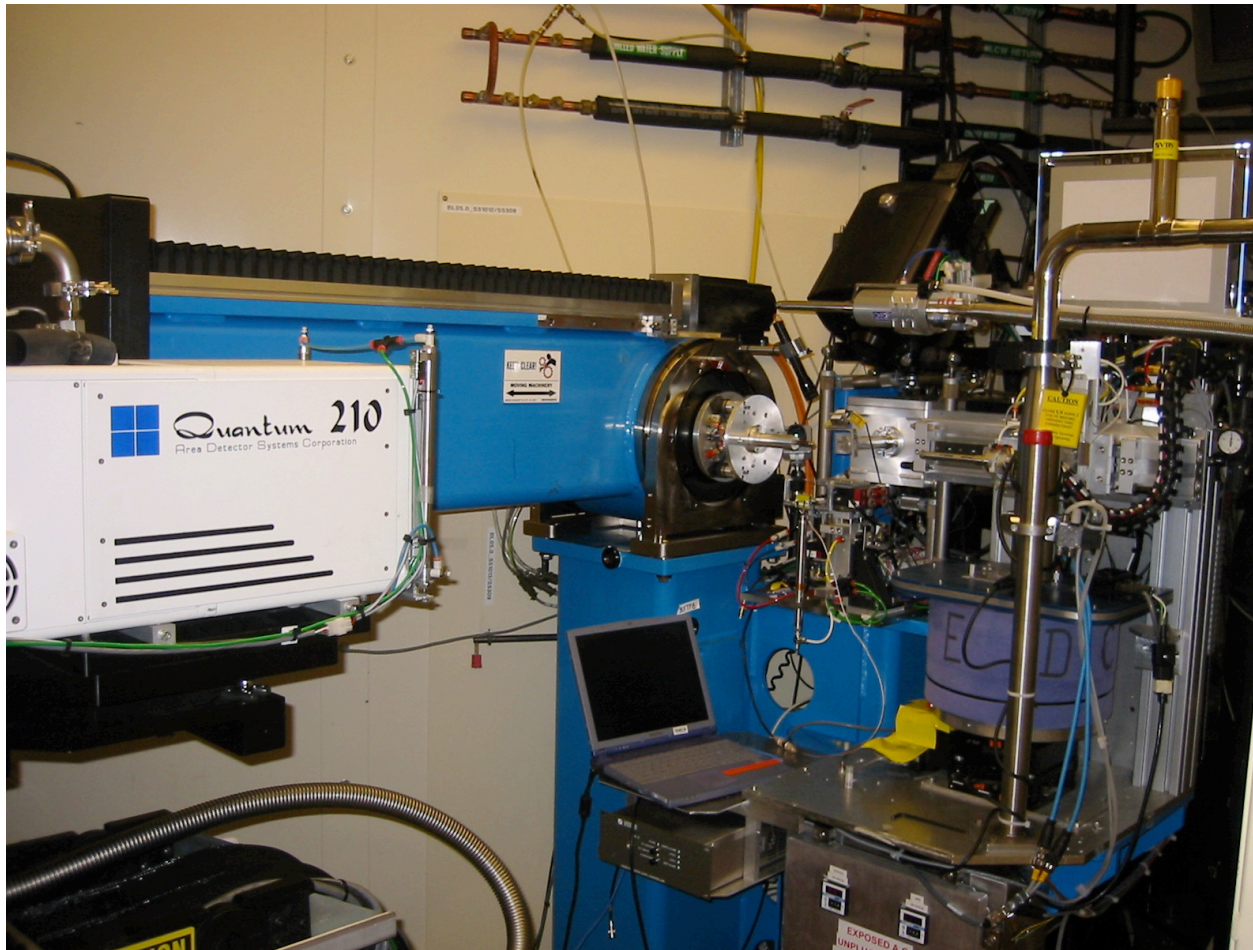


- History
 - Online since 1997. Various upgrades since have made 5.0.2 the 'hottest' PX beam line in the ALS
- Source:
 - 1.96 Tesla, 56 pole, 11.5 cm period permanent magnet wiggler
 - 5.0.2 takes the central 1.5 mrad from the emission fan. Both side-stations (5.0.1 & 5.0.3) take the 2.7 mrad left and right tail.
- Optics:
 - 5.0.2: Cylindrical M1 mirror, flat double xtal mono (LN2 cooled), toroidal M2 mirror on hexapod
 - 5.0.1/5.0.3: Cylindrical M1 mirror, single crystal mono
- 5.0.2: MAD; 5.0.1: Se-SAD; 5.0.3: 1 Å
 - 5.0.3 will be shifted to the Se-HREM in due course
 - Again

Sector 5



Endstations



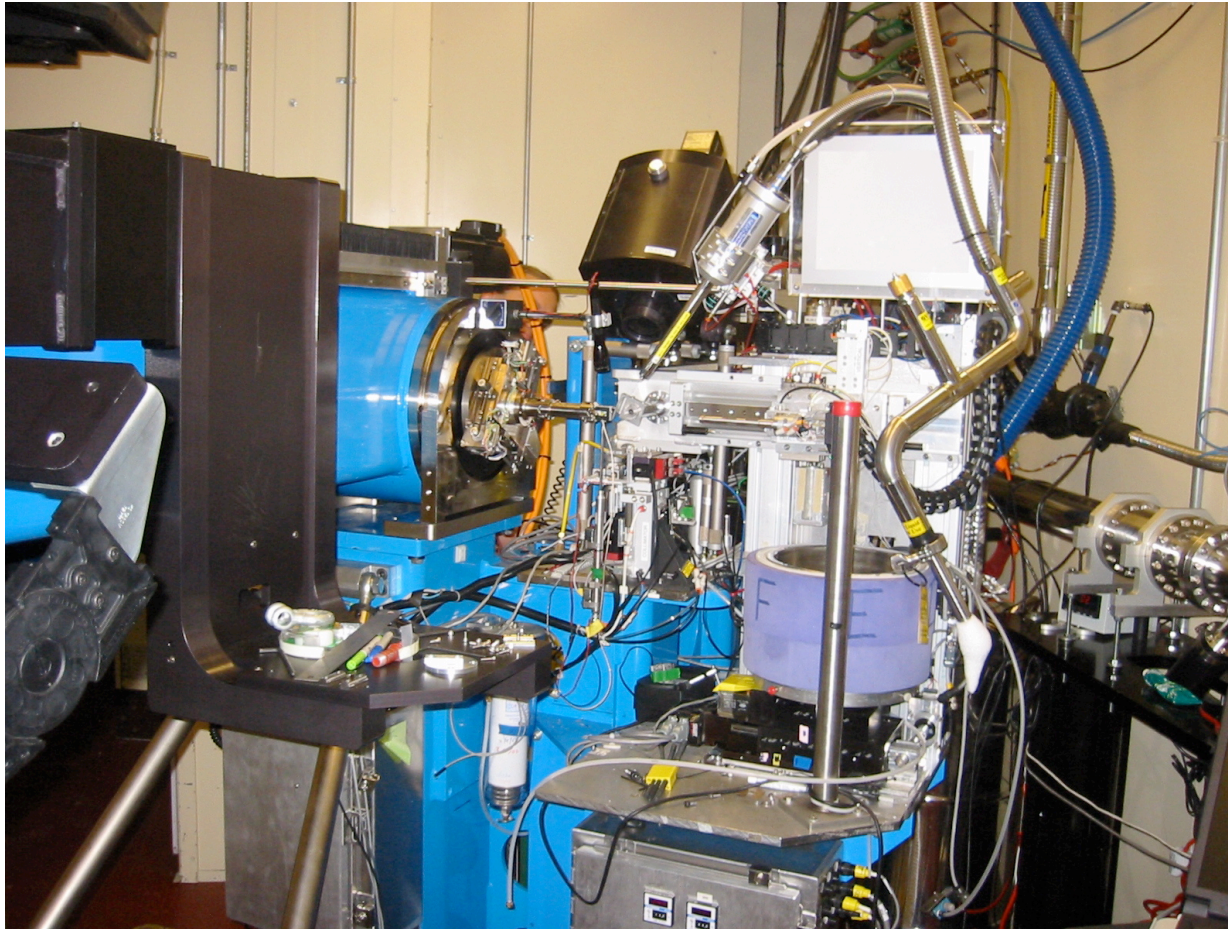
5.0.1:

Q210

Automounter

Two theta arm
(up to 13°)

Endstations



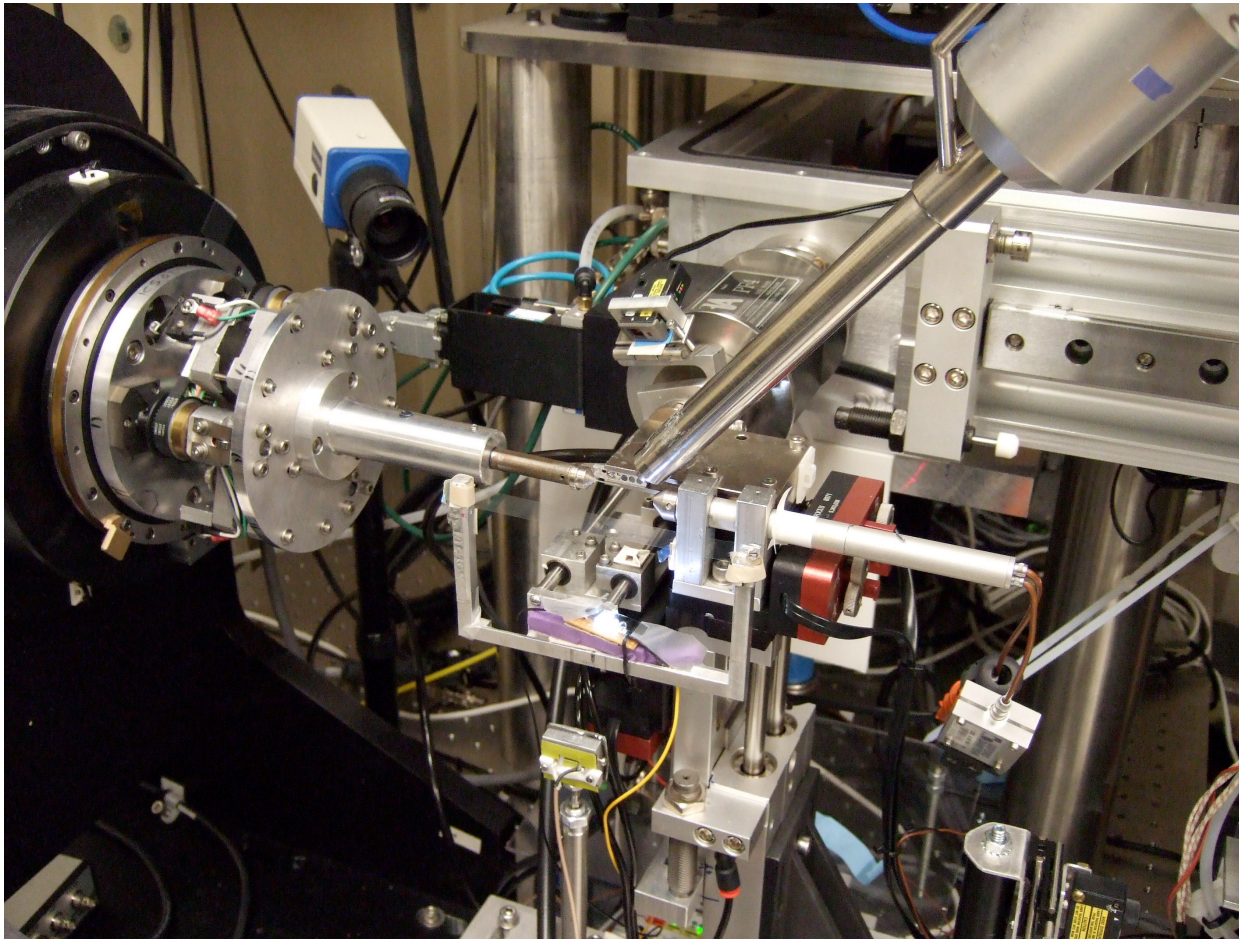
5.0.3:

Q315R

Automounter

Two theta arm
(up to 13°)

Endstations



5.0.2:

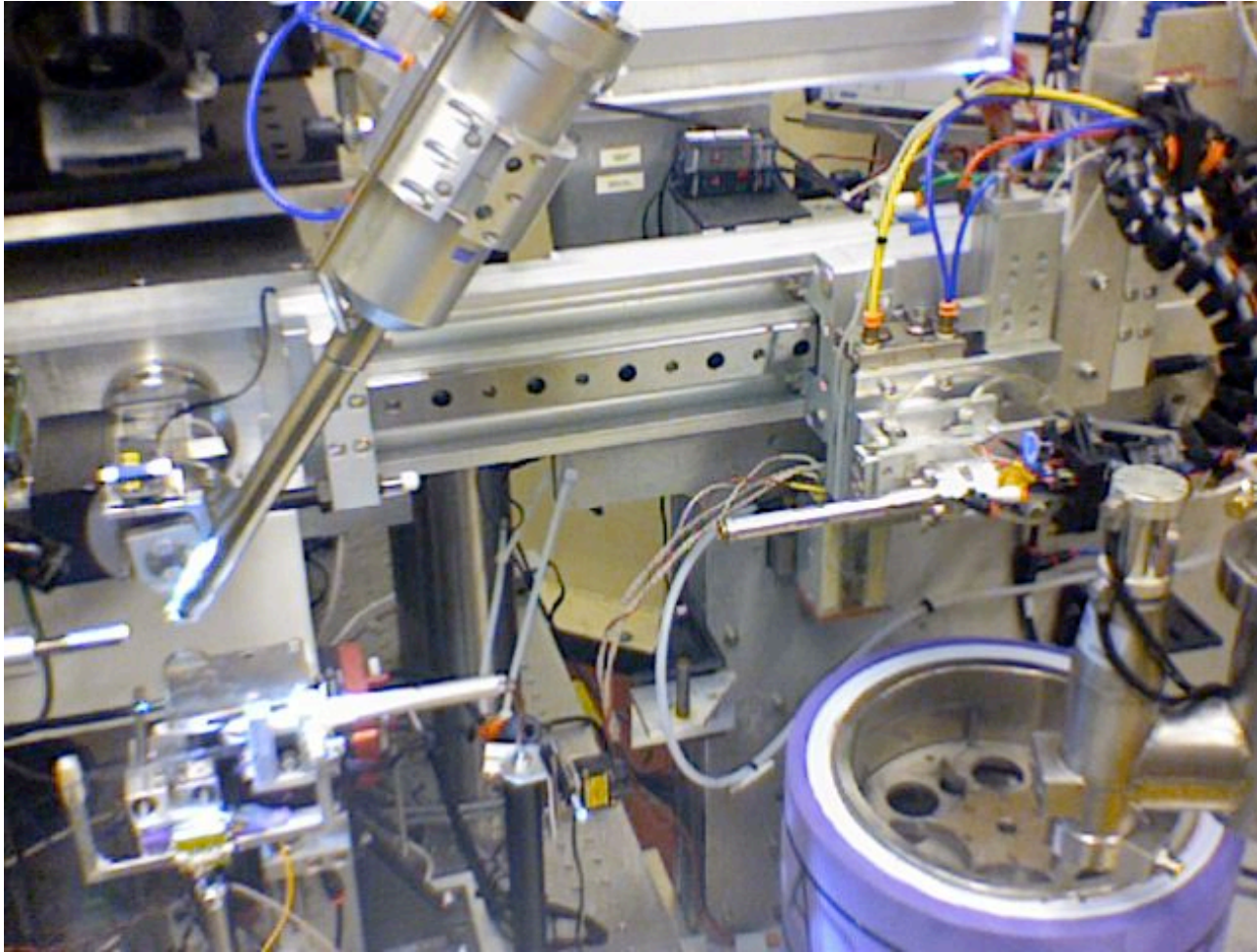
Q315

Automounter

Tuneable

5.5keV - 15.5keV

Automounter



BERKELEY CENTER FOR STRUCTURAL BIOLOGY

PHYSICAL BIOSCIENCES DIVISION

Automounter



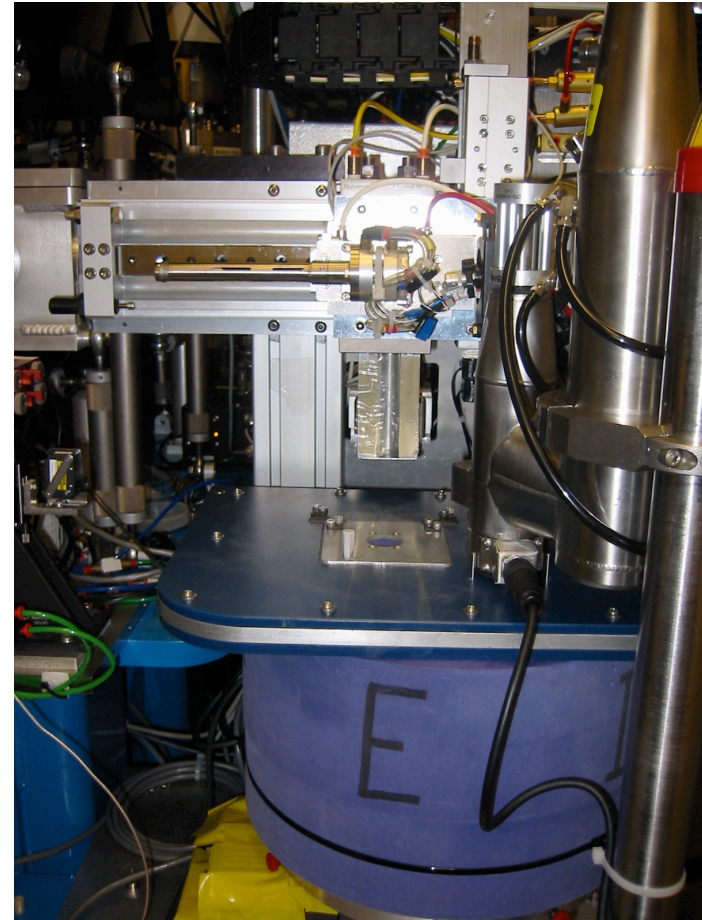
- The LN2 house supply is not stable enough to reliably fill the robot sample dewars
 - Main issues are variable and too high pressure
- A fill system (phase separator and 124 L storage tank) provides LN2 at a constant pressure
 - Over 10 hours of LN2 in case house supply is down



Automounter



- On 5.0.3, the robot is used 100% by the PRT members (approx 60% of the beam time)
- On 5.0.1 and 5.0.2 these number are a bit lower (say 50%)
 - All PRT members use the robot
- Typically, Industrial users handle 2 dewars over 24 hours. This is approximately 160 crystals.
- On a weekly basis, over all sector 5 beamlines, 800 crystals are mounted by the robot.



Efficient screening



- Screening crystals manually is a tedious job, even when using the robot.
- One needs to be in synch with what is on the gonio, and what one writes in ones notes.
- Mistakes are easily made, especially in the small hours.

Efficient screening



- Upload spreadsheet into database
- Tell BOS which puck is in which dewar position
- Generate a queue of crystals
 - Crystal is mounted
 - Wait for user centering
 - Take 0° and 90° shot
 - Pause for user input (shall we collect immediately?)
 - Unmount crystal and mount the next one
- Diffraction patterns can be manually classified

Efficient screening



- The manual evaluation results are updated in the database

Manual Evaluation

Puck = C : Pin = 13 : XID = -1 : Name =

<input type="checkbox"/> Salt	Mosaicity low medium high	Diffraction none Weak Medium Strong	SpotShape Clean Streaked Split Satellite
<input type="checkbox"/> Ice			
<input type="checkbox"/> Collectable			
<input type="checkbox"/> Anisotropic			
Resolution 1.0	Add Evaluation		

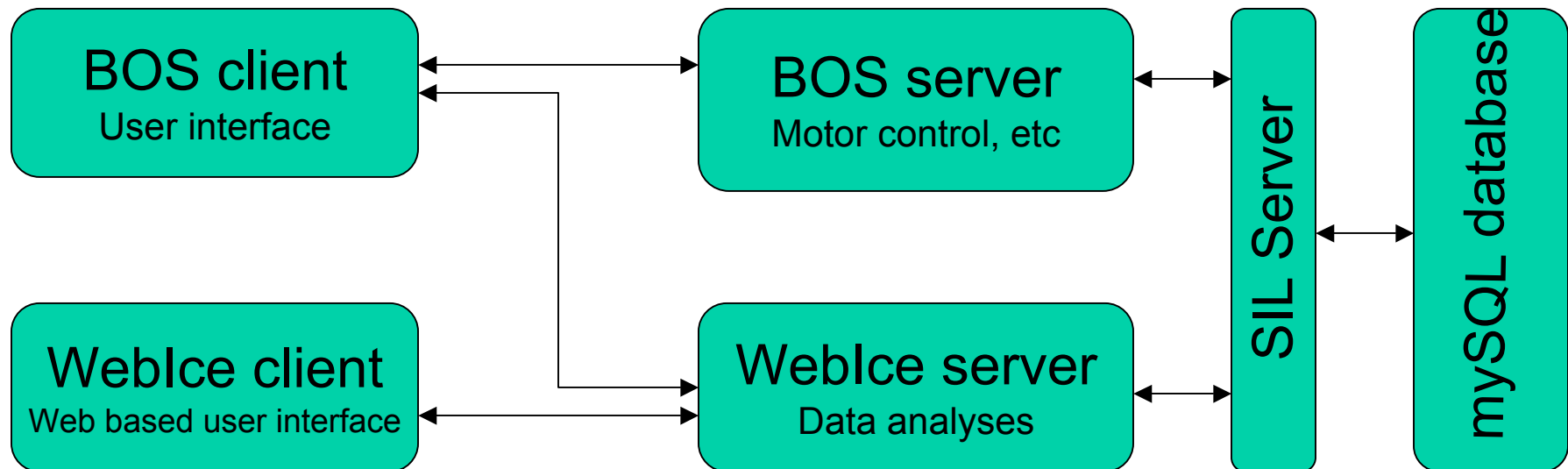
Task Queue Control

Start	Stop	Skip	Redo	Pause	Resume
-------	------	------	------	-------	--------

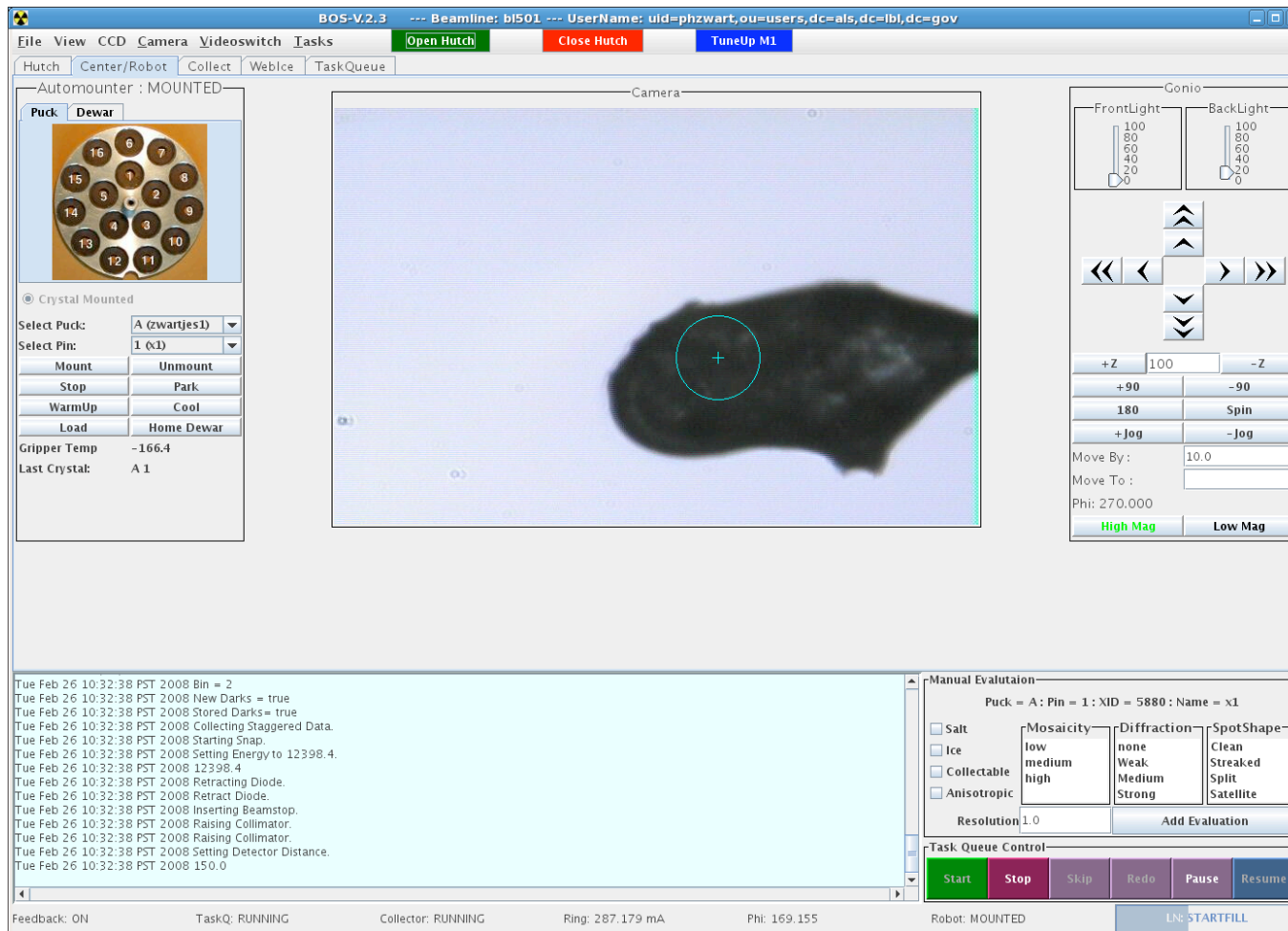
Automated evaluation



- The beamline operating software interfaces with a WebIce server for automated crystal analyses
 - Labelit autoindexing, strategy, ice rings, resolution
 - Results (including jpeg of xtal on gonio and jpegs of diffraction images) are stored in a mySQL database



The interface



The interface



BOS-V.2.3 --- Beamline: bl501 --- UserName: uid=phzward,ou=users,dc=als,dc=lbl,dc=gov

File View CCD Camera Videoswitch Tasks **Open Hutch** **Close Hutch** **TuneUp M1**

Hutch Center/Robot Collect Webce TaskQueue

SIL Configuration

Current SIL: 78

My SILs (id name) 78 zwardjes1

New Puck Delete Puck

New SIL Delete SIL

Upload SIL Download SIL

Webce Evaluations

name	CrystalID	runID	Images	Comment	BraggSpots	IceRings	Score	Resoluti...	Rmsr	Mosaicity	BravaisLattice	UnitCell	SystemWarning
x1	5880	20320	x1_A1_001.img x1_A1_002.img		12	5	0.0						No_indexing_Solution: Too few unimodal Bragg spots (16) in image 1
x2	5881	20321	x2_A2_001.img x2_A2_002.img		990 963	1 0	0.81...	1.56	0.061	0.200000	P4,P422	79.91 79.91 38.48 90.00 90.00 90.00	
x3	5882	20325	x3_A3_001.img x3_A3_002.img		1014 1308	1 0	0.81...	1.43	0.067	0.200000	P4,P422	79.94 79.94 38.49 90.00 90.00 90.00	

Tue Feb 26 11:33:52 PST 2008 Warming Gripper for 10 secs.
 Tue Feb 26 11:33:52 PST 2008 Finished Scan.
 Tue Feb 26 11:33:53 PST 2008 Finished.
 Tue Feb 26 11:33:54 PST 2008 Starting.
 Tue Feb 26 11:33:54 PST 2008 Collecting Image : /data/dcsuser/mintest3/x3/x3_1_074
 Tue Feb 26 11:33:54 PST 2008 Requesting Scan from Beamline.
 Tue Feb 26 11:33:54 PST 2008 startPhi = 73.0 deltaPhi = 1.0 exposure = 1.0
 Tue Feb 26 11:33:57 PST 2008 Finished Scan.
 Tue Feb 26 11:33:58 PST 2008 Finished.
 Tue Feb 26 11:33:58 PST 2008 Starting.
 Tue Feb 26 11:33:58 PST 2008 Collecting Image : /data/dcsuser/mintest3/x3/x3_1_075
 Tue Feb 26 11:33:59 PST 2008 Requesting Scan from Beamline.
 Tue Feb 26 11:33:59 PST 2008 startPhi = 74.0 deltaPhi = 1.0 exposure = 1.0
 Tue Feb 26 11:34:02 PST 2008 Finished Scan.

Manual Evaluation

Puck = A: Pin = 3: XID = 5882: Name = x3

☐ Salt ☐ Ice ☐ Collectable ☐ Anisotropic

Mosaicity: low medium high

Diffraction: none Weak Medium Strong

SpotShape: Clean Streaked Split Satellite

Resolution: 1.0

Add Evaluation

Task Queue Control

Start **Stop** **Skip** **Redo** **Pause** **Resume**

Feedback: ON TaskQ: RUNNING Collector: RUNNING Ring: 260.292 mA Phi: 74.909 Robot: QUEUEPIN **LN: STARTFILL**

The interface



BOS-V.2.3 --- Beamline: bl501 --- UserName: uid=phzward,ou=users,dc=als,dc=lbl,dc=gov

File View CCD Camera Videoswitch Tasks **Open Hutch** **Close Hutch** **TuneUp M1**

Hutch Center/Robot Collect Webce TaskQueue

SIL Configuration

Current SIL: 78

My SILs (id name) 78 zwardjes1

New Puck Delete Puck

New SIL Delete SIL

Upload SIL Download SIL

Webce Evaluations

name	CrystalID	runID	Images	Comment	BraggSpots	IceRings	Score	Resoluti...	Rmsr	Mosaicity	BravaisLattice	UnitCell	SystemWarning
x1	5880	20320	x1_A1_001.img x1_A1_002.img		12	5	0.0						No indexing solution: Too few unimodal Bragg spots (16) in image 1
x2	5881	20321	x2_A2_001.img x2_A2_002.img		990 963	1 0	0.81...	1.56	0.061	0.200000	P4,P422	79.91 79.91 38.48 90.00 90.00 90.00	
x3	5882	20325	x3_A3_001.img x3_A3_002.img		1014 1308	1 0	0.81...	1.43	0.067	0.200000	P4,P422	79.94 79.94 38.49 90.00 90.00 90.00	

Tue Feb 26 11:33:52 PST 2008 Warming Gripper for 10 secs.
 Tue Feb 26 11:33:52 PST 2008 Finished Scan.
 Tue Feb 26 11:33:53 PST 2008 Finished.
 Tue Feb 26 11:33:54 PST 2008 Starting.
 Tue Feb 26 11:33:54 PST 2008 Collecting Image : /data/dcsuser/mintest3/x3/x3_1_074
 Tue Feb 26 11:33:54 PST 2008 Requesting Scan from Beamline.
 Tue Feb 26 11:33:54 PST 2008 startPhi = 73.0 deltaPhi = 1.0 exposure = 1.0
 Tue Feb 26 11:33:57 PST 2008 Finished Scan.
 Tue Feb 26 11:33:58 PST 2008 Finished.
 Tue Feb 26 11:33:58 PST 2008 Starting.
 Tue Feb 26 11:33:58 PST 2008 Collecting Image : /data/dcsuser/mintest3/x3/x3_1_075
 Tue Feb 26 11:33:59 PST 2008 Requesting Scan from Beamline.
 Tue Feb 26 11:33:59 PST 2008 startPhi = 74.0 deltaPhi = 1.0 exposure = 1.0
 Tue Feb 26 11:34:02 PST 2008 Finished Scan.

Manual Evaluation

Puck = A: Pin = 3: XID = 5882: Name = x3

☐ Salt ☐ Ice ☐ Collectable ☐ Anisotropic

Mosaicity: low medium high

Diffraction: none Weak Medium Strong

SpotShape: Clean Streaked Split Satellite

Resolution: 1.0

Add Evaluation

Task Queue Control

Start **Stop** **Skip** **Redo** **Pause** **Resume**

Feedback: ON TaskQ: RUNNING Collector: RUNNING Ring: 260.292 mA Phi: 74.909 Robot: QUEUEPIN LN: STARTFILL

User support



- Sector 5: 1.5 scientist, 2 SEA's
- Sector 8: 1.5 scientist, 1SEA, 1 RA
- From 1600 to 2400 1 SEA
- From 2200 to 0600 1 SEA
- Weekend: from 0900 to 1300 1 SEA

User support



- With various people picking up where others left, efficient communication is vital for a smooth operation
- Our beamline blog is very useful in this respect and serves as a 'long term memory' of known issues and solutions.

The screenshot displays the 'BCSB Beamline Status Log' website. The page has a yellow header with the title 'BCSB Beamline Status Log'. Below the header, there are two blog entries. The first entry is dated '26/03' and titled 'Failure to take snapshots on BL5.0.2'. It includes a category '5.0.2', author 'JMD', and a detailed description of a problem with taking snapshots, a solution involving the motor driver, and a final update. The second entry is dated '26/03' and titled '5.0.2 beam intensity low'. It includes a category 'General', author 'DB', and a description of a low beam intensity issue. On the right side of the page, there are navigation links (Today, Archives, Admin), categories (All, 5.0.1, 5.0.2, 5.0.3, 8.2.1, 8.2.2, General), a search bar, a login section (logged in as phzwart), and a list of links to various resources like nucleuscms.org, docs.nucleuscms.org, forum.nucleuscms.org, wiki.nucleuscms.org, skins.nucleuscms.org, plugins.nucleuscms.org, and dev.nucleuscms.org. The Nucleus logo is also visible at the bottom right of the page.

Phenix



- Phenix aims to automate crystallography
 - Assumed is the presence of reduced data
 - It perform all tasks up to validation
 - Easy to use command line:

```
phenix.autosol 40 Se seq.txt
```

```
phenix.automr model.pdb data.mtz
```

```
phenix.refine data.sca model.pdb
```

```
phenix.xtriage data.mtz
```


xtriage



- Perform a number of basic sanity checks on the data
- Has extensive twinning analyses
 - Almost all twin tests known to mankind are performed
- Tells you what is going on!
 - Not: $\langle |L| \rangle = 0.43$
 - But: you have twin laws, intensity statistics are abnormal, your data might be twinned.
- Informs you if point group of data is too low, or when unit cell might be too big

Other twin related features



- The tools presented here are part of the *phenix* suite
 - <http://www.phenix-online.org>
- Key applications for twinning
 - *phenix.xtriage* : Detection of twinning
 - *phenix.refine* : Refinement of twinned data
 - *iotbx.explore_metric_symmetry* : understanding relations between space groups

Detection of twinning



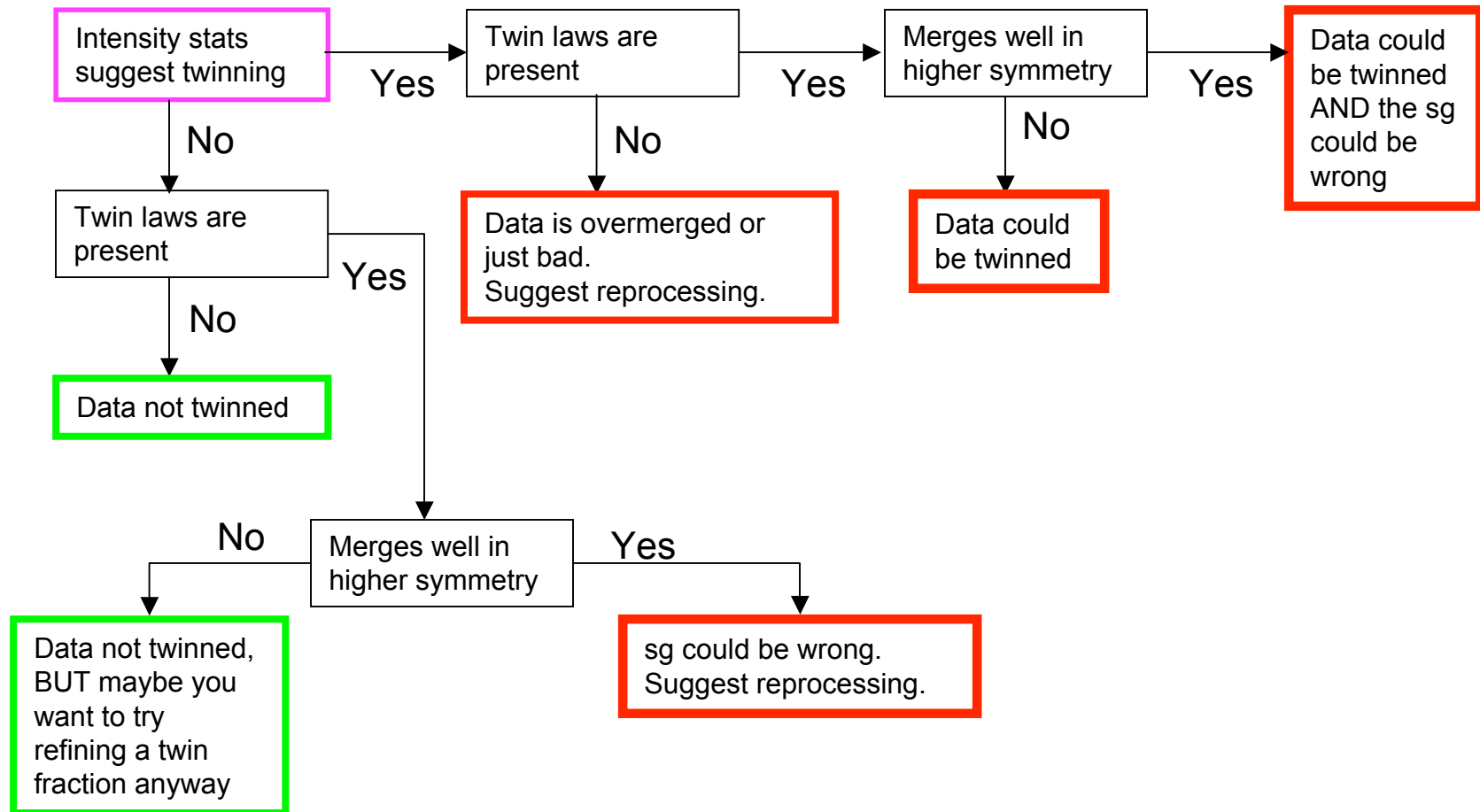
- Twinning can be spotted by inspecting intensity statistics
 - Values of intensity statistics are known for untwinned data
 - $\langle |I|^2 \rangle / \langle |I| \rangle^2$ Intensity ratio
 - $\langle F \rangle^2 / \langle F^2 \rangle$ Amplitude ratio
 - $\langle |E^2 - 1| \rangle$
 - $\langle |L| \rangle$ Local intensity statistic
 - Cumulative intensity distribution (NZ plot)
- All these statistics are very sensitive to the quality of the data
 - Data to be used in intensity statistics is cut at a resolution shell where 85% of the data still has $I/\sigma_I > 3$ (xtriage default)
 - This eliminates noisy shells and ‘stabilizes’ intensity statistics
- What are good values though?
 - Over 5000 data sets of non-twinned data build up ‘crystallographic intuition’

Twin laws



- Determination of twin laws
 - From first principles given your uc and sg
 - No twin law will be overlooked
 - Not all of the available twinning detection tools are as thorough as needed
 - If lookup tables are used, pseudo merohedral twinning can be missed
- PDB analyses: 36% of structures has at least 1 possible twin law
 - 50.9% merohedral; 48.2% pseudo merohedral; 0.9% both
 - 27% of cases with twin laws has intensity statistics that warrant further investigation on whether or not the data is twinned
 - **10% of whole PDB(!)**

Twinning



Other xtriage features



- Other useful statistics given by *phenix.xtriage*
 - Cumulative intensity distribution
 - R vs R statistic
 - This is what you need to demonstrate twinning in the presence of pseudo symmetry
 - Lebedev, Vagin, Murshudov. *Acta Cryst.* (2006). **D62**, 83-95
 - Britton plot
 - H-test
 - Likelihood based twin fraction estimate
 - Very much like a Murray-Rust plot actually
 - CCP4 style plots

Refinement of twinned data



- The twin target function used in *phenix.refine* is similar to the one used in CNS and refines against the twinned amplitudes:

$$\sum_{\mathbf{h1}} w_{\mathbf{h1}} \left(F_{\mathbf{h1},obs} - \sqrt{(1-\alpha)F_{\mathbf{h1},calc}^2 + (\alpha)F_{\mathbf{h2},calc}^2} \right)^2$$

$$\mathbf{F}_{\mathbf{h1},calc} = f(k_{overall}, B_{overall}) (\mathbf{F}_{\mathbf{h1},atoms} + f(k_{sol}, B_{sol}) \mathbf{F}_{\mathbf{h1},bulk})$$

- Twin fraction and overall and bulk solvent scale parameters are optimized using robust derivative free optimizer
 - This is done before positional and ADP refinement in *phenix.refine*

Refinement of twinned data

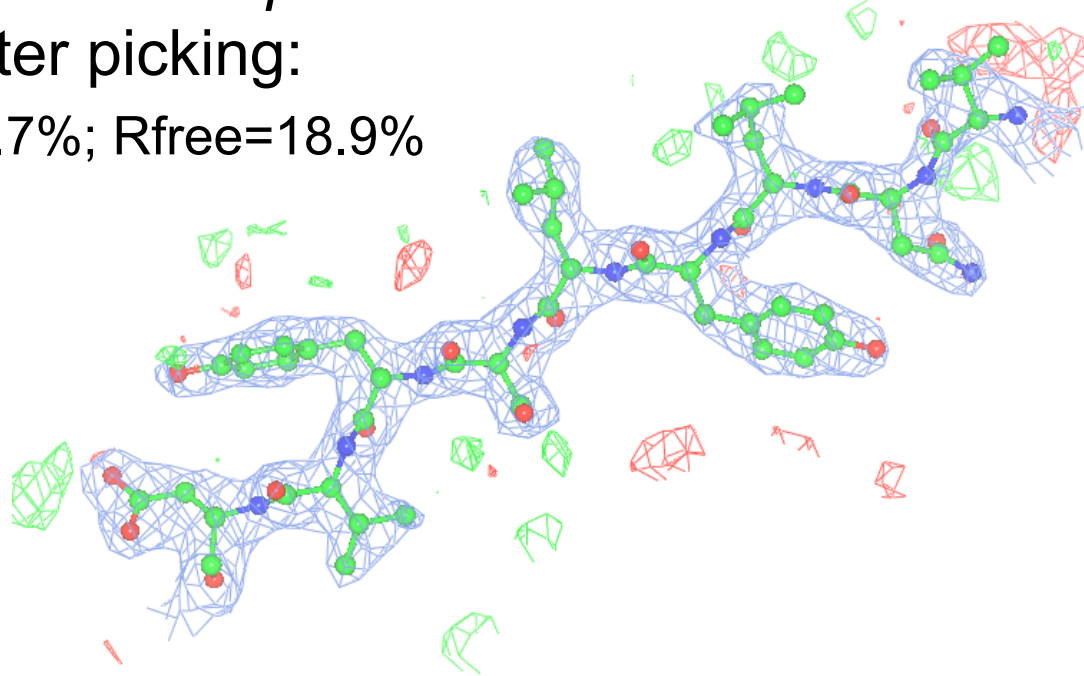


- The twin target in *phenix.refine* is (almost) like any other target. It allows for refinement of
 - Rigid body refinement
 - Refinement by simulated annealing
 - Group B factor refinement
 - Occupancy refinement
 - f' and f'' refinement
 - Refinement of TLS parameters
 - Refinement of anisotropic parameters
 - Refinement of 'inter-atomic scatterers'
 - Modeling bond electrons

Example: Porin



- CNS 'standard'; after refinement of xyz, B and water picking:
 - Rwork=14.6%; Rfree=18.7%
- Same model used in *phenix.refine*. After refinement of xyz, B + water picking:
 - Rwork=14.7%; Rfree=18.9%

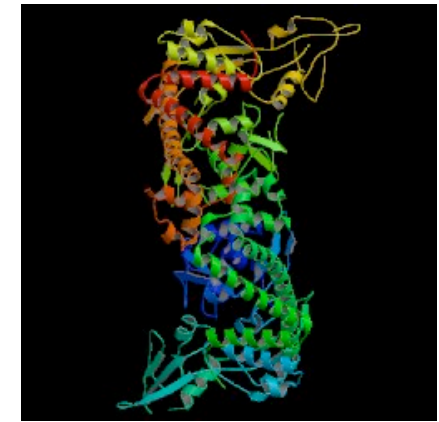


TLS and twinned data

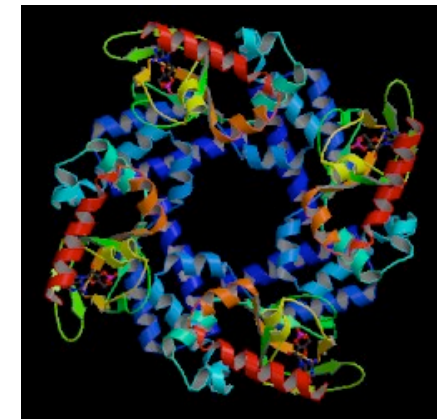


- Twinned structures quite often have more than a single copy in the ASU
 - TLS can be useful in those cases
 - Examples of effects of impact of TLS on R-values

PDBID	Without TLS	With TLS	Δ	α
2NOV	22.1 / 27.2	20.1 / 25.9	-2 / -1.3	0.38
2QA0	20.7 / 27.8	20.5 / 25.8	-0.2 / -2	0.43
1Q43	17.1 / 23.8	15.1 / 21.1	-2 / -2.7	0.09
1Q3E	15.9 / 20.5	15.3 / 19.9	-0.6 / -0.6	0.12



2NOV

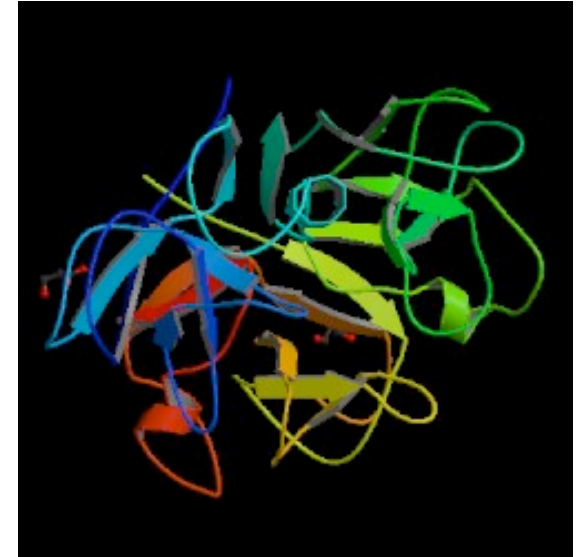


2Q0A

Anomalous twinned data



- gpD (capsid stabilising protein of bacteriophage lambda) Se-MAD data (courtesy of Z. Dauter)
- (45.5 68.5 45.5 90 104.5 90), $P2_1$
 - Possible twin law: $(-l, -k, -h)$
- Intensity statistics suggest twinning
 - $\langle |L| \rangle$: 0.387 (Z-score: 10)
 - $\langle |E^2 - 1| \rangle$: 0.573 (Z-score: 8)
- Final model available from PDB, originally solved by MAD methods
 - Yang *et al*, Acta Cryst. (2000). **D56**, 959-964
- Re-refine structure against twinned MAD data



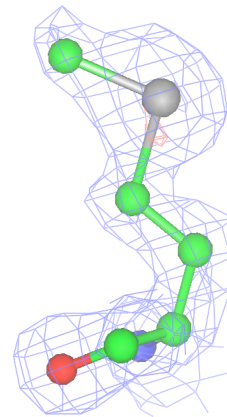
1c5e

Anomalous twinned data

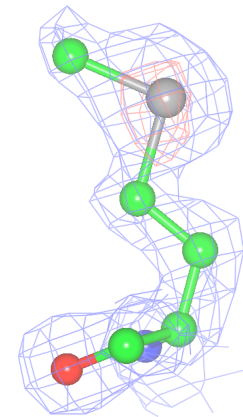


- No model updates
- Refine f' and f'' with all other model parameters:

	f'	f''	α
inflection	-6.2	3.0	0.32
peak	-5.2	5.3	0.33
remote	-4.3	3.6	0.32



Refine f' and f''
Peak data



$f'=0$, $f''=0$

phenix.xtriage usage



- Designed to be easy to use, automatic, configurable for experts and have easy to interpret output

`phenix.xtriage mydata.sca`

Statistics independent of twin laws

- $\langle I^2 \rangle / \langle I \rangle^2$: 1.587
- $\langle F^2 \rangle / \langle F \rangle^2$: 0.871
- $\langle |E|^2 - 1 \rangle$: 0.573
- $\langle |L| \rangle$, $\langle L^2 \rangle$: 0.387, 0.212
Multivariate Z score L-test: 10.237

Statistics depending on twin laws

Operator	type	R obs.	Britton alpha	H alpha	ML alpha
-l, -k, -h	PM	0.164	0.335	0.328	0.311

The results of the L-test indicate that the intensity statistics are significantly different than is expected from good to reasonable, untwinned data.

As there are twin laws possible given the crystal symmetry, twinning could be the reason for the departure of the intensity statistics from normality.

It might be worthwhile carrying out refinement with a twin specific target function.

phenix.refine usage



- Designed to be easy to use, automatic, configurable for experts and have easy to interpret output

```
phenix.refine data.mtz model.pdb \  
twin_law="-h-k,k,-1"
```

```
model_refine_001.log
```

```
model_refine_001.pdb
```

```
model_refine_map_coefs_001.mtz
```

```
model_refine_002.def
```


Acknowledgements



BCSB

Paul Adams

Simon Morton
Corie Ralston

Banumathi Sankaran

Nathan Smith
Azer Daut
Jeff Dickert
Anthony Rozales
Diane Briant

Yun Zhou
John Taylor

Users

Virginia Rath (rsc)
Gyorgy Snell (TakedaSD)
Glenn Spraggon (GNF)
Dan Knighton (Pfizer)
Jeff Abramson (UCLA)

Funding :

Amgen

Vertex

LANL/TBSGX

UCSF

Gilead

Pfizer

FHCRC

Genentech

Celgene

Roche

GNF

TakedaSD

Acknowledgements



LBNL

Ralf Grosse-Kunstleve
Pavel Afonine
Nick Sauter
Nigel Moriarty
Paul Adams
(Nat Echols)

LANL

Tom Terwilliger
Li-Wei Hung

Cambridge University

Randy Read
Airly McCoy
Laurant Storoni
Gabor Bunkoczi
Robert Oeffner

Duke University

Jane Richardson
Dave Richardson
Ian Davis
Vincent Chen
Jeff Headd

Texas A&M

Tom Ioerger
Eric McKee

Funding & feedback:

phenix industrial consortium

Boehringer Ingelheim Pharmaceuticals, Inc.
Glaxo-Smith-Kline
Johnson and Johnson
Novartis Pharmaceuticals Corp.
Plexxikon Inc.
Wyeth Ayerst Research

Funding:

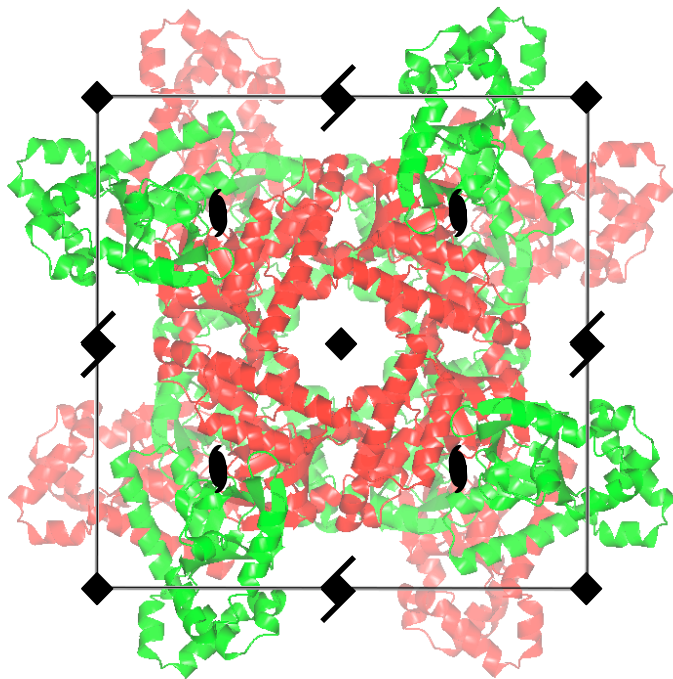
NIH

LBL

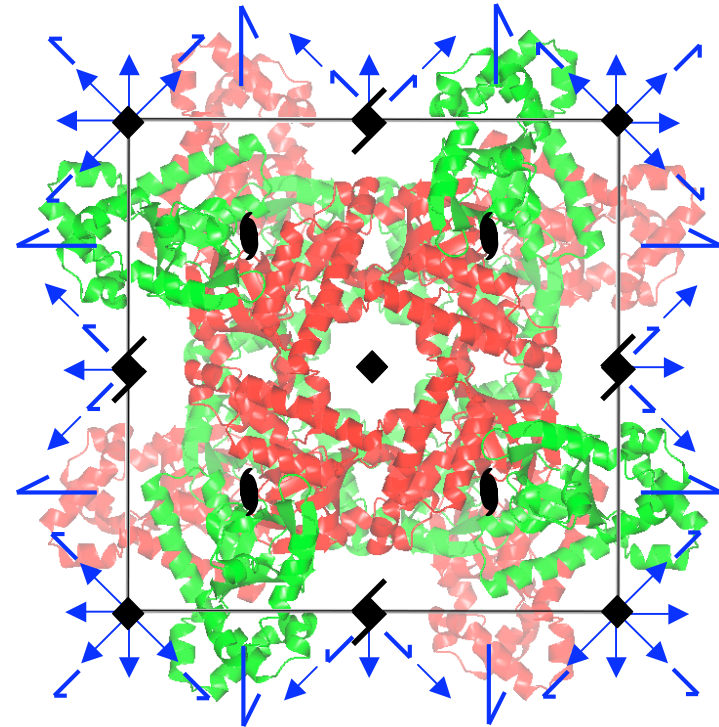
Feedback:

phenixbb

Thank you!



I4



I422